

# UNCLASSIFIED

AD NUMBER
AD861129
NEW LIMITATION CHANGE
TO Approved for public release, distribution unlimited
FROM Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational use; Oct 1969. Other requests shall be referred to Rome Air Development Center, Griffiss AFB, NY 13441-5700.
AUTHORITY
Per RADC ltr, 30 Jan 1973

THIS PAGE IS UNCLASSIFIED

7  
AD861129

RADC-TR-69-308  
Final Technical Report  
October 1969

COMPILATION OF DOD SCIENTIFIC AND TECHNICAL  
INTELLIGENCE THESAURUS  
System Development Corporation

This document may be further  
distributed by any holder only  
with specific prior approval of  
DIA (OASST-1), Wash. D. C.  
20315.

D D C  
RECEIVED  
NOV 10 1969  
DISSEMINATION

Home Air Development Center  
Air Force Systems Command  
Griffis Air Force Base, New York

Best Available Copy

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related government procurement operation, the government assumes no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded, by implication or otherwise, as in any manner endorsing the holder or any other person or corporation, or conveying any right or authority to manufacture, use, or sell any patented invention that may in any way be covered by the said drawings, specifications, or other data.

Do not return this copy. Retain or destroy.

**Best Available Copy**

**COMPILATION OF DOD SCIENTIFIC AND TECHNICAL  
INTELLIGENCE THESAURUS**

**John A. Dovel, Jr.**

**System Development Corporation**

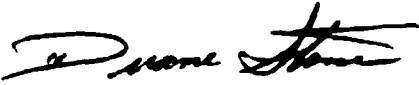
**This document may be further  
distributed by any holder only  
with specific prior approval of  
DIA (DIAST-1), Wash. D. C.  
20315.**

# FOREWORD

This technical report was prepared by Mr. John A. Dovel, Jr. of System Development Corporation, 5720 Columbia Pike, Falls Church, Virginia 22041 under Contract F30602-68-C-0252, Project 9117. The Rome Air Development Center project engineer was Mr. Duane Stone, EMBIH. Contractor's report number is TM-WD-(L)-321/000/00.

This technical report has been reviewed and is approved.

Approved:

DUANE L. STONE   
Project Engineer  
Intelligence Data Handling Branch

Approved:

  
A. E. STOLL, Colonel, USAF  
Chief, Intel & Info Processing Div.

FOR THE COMMANDER:

  
IRVING J. GABELMAN  
Chief, Plans Office

# ABSTRACT

A thesaurus, consisting of 22,540 terms, was compiled for use in indexing and retrieving scientific and technical intelligence information. The thesaurus (covering activities monitored by Defense Intelligence Agency) can be printed by computer in three forms - subject structured display, permuted display and alphabetical display. The philosophy and procedures governing thesaurus maintenance were also studied.

TABLE OF CONTENTS

	<u>Page</u>
Thesaurus Compilation .....	1
Appendix I Study and Analysis of Maintenance Procedures in CIRC Operations .....	3
Appendix II Comments on the CIRC Thesaurus .....	23

## THESAURUS COMPILATION

System Development Corporation had contractual relations with various members of the intelligence community which resulted in developing thesauri for scientific and technical intelligence activities. This work indicated a sufficient commonality of interests between the community members to warrant development of an integrated S&T thesaurus to satisfy the needs of all DOD military S&T intelligence activities monitored by the Defense Intelligence Agency (DIA).

The DOD S&T Intelligence Thesaurus is a compilation of microthesauri produced for the Army Missile Intelligence Directorate (MID), MICOM, Huntsville, Alabama; the Army Medical Intelligence Office (MIO), Washington, D.C.; the Navy Scientific and Technical Intelligence Center (STIC), Washington D.C.; and the Directorate for Scientific and Technical Intelligence, DIA (DIAST-2C). It includes as a base and model, the CIRC Thesaurus which had been developed at the Foreign Technology Division (FTD), AFSC, Wright-Patterson AFB, Dayton, Ohio. During the course of this contract FTD personnel consolidated the CIRC thesaurus with the Vocabulary of Intelligence Concept Expressions (VOICE), developed for the Army Foreign Science and Technology Center (FSTC), Washington, D.C. These changes were forwarded to the Falls Church Office of System Development Corporation (SDC) as they became available and integrated into the final thesaurus.

Upon receipt of the CIRC vocabulary on magnetic tapes from FTD, the data was converted to a deck of cards, listed and compared with titles in subject files at each agency and appropriate interest profiles. The results of this comparison was the basis for a second subset or deck of terms for a particular microthesaurus. These segments were then scrutinized in order to determine the most appropriate location for additional terminology. This intensive study revealed situations with regard to the selection and organization of certain CIRC terms that required considerable adjustment before terms for the microthesauri could be added. Duplicate cards in the microthesaurus decks and the 56 subject areas of CIRC were used to assure consistency with published precedents. Considerable effort was made to temporize on changes in the interest of continuity.

The DOD S&T Intelligence Thesaurus is divided into three parts; a subject-structured display, in which the terms are hierarchically related within 56 broad subject areas; a permuted display,



in which the single and multi-word terms are arranged by computer in the order of each word that appears within the terms; and an alphabetical display in which the terms are listed alphabetically with cross references among subject-related or hierarchically related terms and to synonyms where applicable, and with scope notes to indicate the intended usage of ambiguous terms. Publication is scheduled for the fall of 1969.

All microthesaurus terms are incorporated into the DOD S&T Intelligence Thesaurus. However, there are slight differences in structuring between the micro-thesauri and the thesaurus to be published by FTD. These differences arise because of program developments at FTD during the contract period. The micro-thesauri listings are being produced at a local facility and utilize a 1968 version of the Dayton computer program. Consequently, they are structured in a format identical to the last published CIRC thesaurus. Input material to the contract thesaurus has been prepared for a 1969 version of the same computer program which provides somewhat increased capabilities for term display.

A statistical comparison of the various entries in the December 1968 CIRC Thesaurus and the DOD S&T Intelligence Thesaurus is shown below.

	December 1968 CIRC Thesaurus	July 1969 DOD S&T Intel Thesaurus
Official Terms (OT)	12,999	15,427
Official Term Synonyms (SY)	1,010	2,071
Official Nomenclature Terms (ONT)	4,425	4,509
Official Nomenclature Term Synonyms (ONT-SY)	302	533
Total entries	18,736	22,540

These figures reveal a net increase of 3,904 entries in the size of the vocabulary. Actually, the total change in the vocabulary is far greater since the consolidation of the micro-thesauri required many term substitutions and revisions, as well as extensive reorganization of the vocabulary in several areas.

## APPENDIX I

### STUDY AND ANALYSIS OF MAINTENANCE PROCEDURES IN CIRC OPERATIONS

#### SECTION I

##### REVIEW OF THE CIRC THESAURUS MAINTENANCE PROCEDURES

###### INTRODUCTION

Pursuant to a task requirement of Contract F30602-68-C-0252 (PR-I-8-4466), a study was made of applicability of the present vocabulary support effort at FTD to an expanded CIRC. Recommendations are made for systematic maintenance of the DOD Scientific and Technical Intelligence Thesaurus.

An overview of the present FTD vocabulary support system is given. A suggested plan is presented for enlarging the scope of present support system procedures to accommodate new users and an extended vocabulary. The recommended operating procedures mirror the present system with few innovations. It is hoped that sufficient detail and references are presented to provide the basis for a primer or handbook for new activities entering the CIRC system.

An investigation was made of a new approach to vocabulary support within the CIRC system. The techniques are not new and can be seen during any demonstration of COLEX or a time-shared management system. Basically, the approach is to bring a direct-access automatic updating capability to the lexicographic function. The project, called AUTOLEX (for automating lexicographic functions), was initiated in order to derive some basic conclusions and recommendations for further work. The project is described in Section II.

###### METHODOLOGY

The review of the present vocabulary support system at FTD was accomplished through direct conversations with the lexicographic group (TDBAC), members of the Request Center (TDBIR), the internal indexing group (TABAC-2). In addition, the experience and knowledge of the SDC Dayton Office was utilized by direct contact and review of pertinent documentation.

It was decided that the AUTOLEX project would best be demonstrated utilizing a CIRC-like data base operated on by the direct access capability of the SDC Time-Shared Data Management System (TDMS). TDMS was chosen over the On-line Retrieval of Bibliographic Information Time-shared (ORBIT) System only because the up-date capability of ORBIT was not complete at the time of the study.

###### CONSTRAINTS

At no time was the goal of this effort considered as license to redesign the entire vocabulary support system. To opt otherwise would generate massive changes in a number of production programs because of the vocabulary dependencies. Rather the aim of this effort focused on finding the least disruptive approach, in terms of resources expended, to integrating the microthesauri into an expanded CIRC while providing the new activities with sufficient means to maintain a current vocabulary.

Only those lexicographic functions which deal directly with the interaction of users and the lexicographic group (TDBAC) were considered for investigation - a decision which affected the detailing of problem areas encountered. Specific information on equipment changes became known prior to publication, but after this investigation had completed data collection phases. Since the study was undertaken while the entire CIRC system was anticipating an equipment change-over from an IBM-7094 to an IBM0360/65, it was assumed that conversion to third-generation equipment would necessitate alterations to the present operational format for vocabulary update and support.

#### PRESENT SYSTEM - OVERVIEW

The operating environment of the lexicographic group is depicted in Figure A1. Communications with the user groups may occur through information specialists at the Request Center or come about directly from the user to the lexicographic group. The indexer to lexicography office channel has always been one of a direct exchange between the lexicographic group and a single source at each indexing unit. The following sections describe those lexicographic functions which relate directly to the transfer of information from the user to the lexicography office.

#### VOCABULARY CHANGE

The CIRC vocabulary is meant to be highly "user oriented". Direct suggestions from indexers and users are vital to insure a current and comprehensive vocabulary. At present, user suggestions for vocabulary change take place through the use of a Term Control Form, FTD-0-87 (see Figures A2 and A3. One copy of this four-part form is kept in suspense at the lexicography office; the remaining copies are forwarded to the lexicography office through quality control points. Final disposition of the request is communicated to the initiator by means of copies of the same form. The form itself is adequate for change requests and, provided the instructions are followed correctly, no further information is required.

A CIRC term control file is maintained within the lexicography office. Figure A4, shows the log-in procedures and duplication check made on all incoming term control forms. The term control file consists of suspense and action subfiles. The action file depicts all dispositions made on requested changes to the vocabulary and is represented by the yellow, or fourth, copy of the form. The suspense file represents those requests not yet processed and consists of copies 1, 2, and 3 of the form. Both files are maintained in alphabetical sequence by term.

Incoming requests are checked against the suspense file for like occurrences originating from another activity and awaiting disposition. When the incoming request is found to be repetitive with a prior request, the new form is clipped to the old one and returned to the suspense file. When file disposition is made on the term in question, all activities requesting the term change are notified.

When the check on the suspense file is negative, the action file is searched. If the alphabetic search proves fruitless, the form copies are placed in proper sequence in the suspense file. When it is found that the term has had prior action, items 6-10 of the third copy are completed from the previous disposition and returned to the latest originator.

31 July 1969

TM-WD-(L)-321/000/00

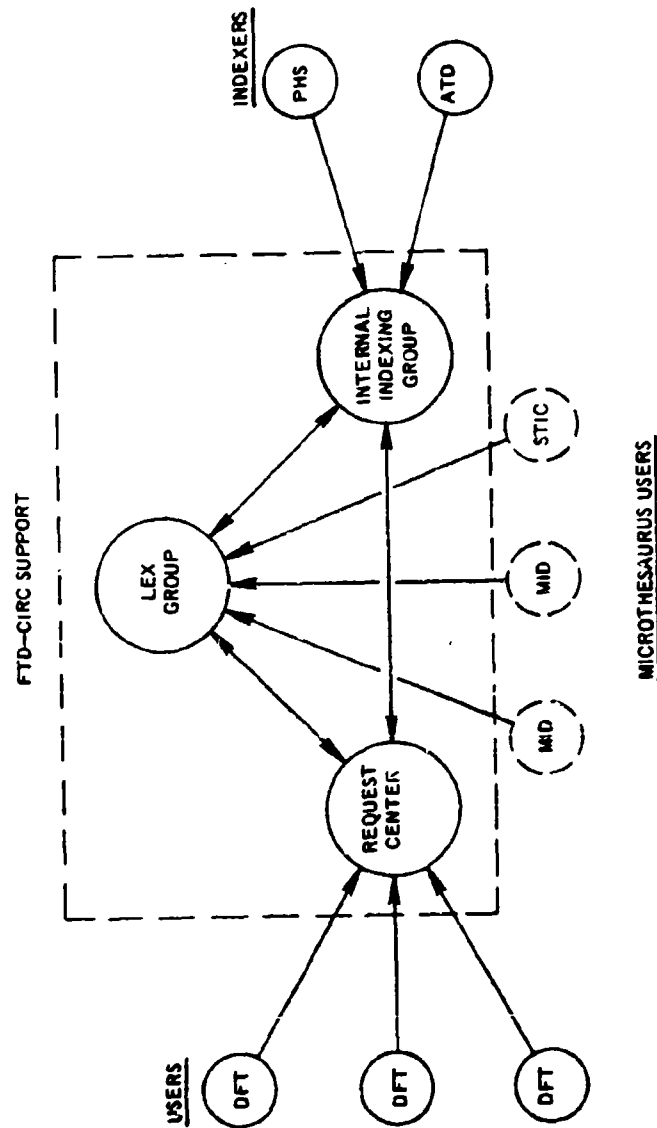


Fig. A1 CIRC Lexicographic Operating Environment

31 July 1969

TM-WD-(L)-321/000/00

CIRC TERM CONTROL		ORIGINATOR (complete items 1-5 ONLY)	DATE
1. NAME/ORGANIZATION		2. TERM	
3. DEFINITION/BROADER TERM			
4. CLASSIFICATION OF TERM <input type="checkbox"/> U <input type="checkbox"/> C <input type="checkbox"/> S <input type="checkbox"/> NS		5. ACTION REQUESTED A. <input type="checkbox"/> ADD B. <input type="checkbox"/> DELETE	
CLASSIFICATION SOURCE/AUTHORITY		C. <input type="checkbox"/> CHANGE TO:	
FOR LEXICOGRAPHIC USE ONLY			
6. ACTION TAKEN:		<input type="checkbox"/> OT OR ONT: <input type="checkbox"/> SY OF: <input type="checkbox"/> VOREF EQUATE TO:	
<input type="checkbox"/> A. ACCEPTED AS			
<input type="checkbox"/> B. MODIFIED AND ACCEPTED AS			
<input type="checkbox"/> C. HELD FOR FUTURE REVIEW			
<input type="checkbox"/> D. RETURNED			
		<input type="checkbox"/> CONCEPT EXPRESSED BY TERM IS TOO BROAD	
		<input type="checkbox"/> CONCEPT EXPRESSED BY TERM IS TOO SPECIFIC	
		<input type="checkbox"/> INSUFFICIENT INFORMATION GIVEN ABOUT TERM	
		<input type="checkbox"/> NO INFORMATION NOT GIVEN	
7. CLASSIFICATION CHANGES:			
FROM: ; TO: ; DATE: ; AUTHORITY:			
FROM: ; TO: ; DATE: ; AUTHORITY:			
8. REMARKS:			
9. NAME AND DATE OF REVIEWER		10. EFFECTIVE DATE	

FTD FORM 0-37 APR 66 REPLACES FTD FORM 0-1 (TEST) NOV 67

Fig. A2 Term Control Form

### INSTRUCTIONS

Consult the CIRC Thesaurus Introduction, for instructions on term construction.

This form is to be used for all proposed additions/deletions/modifications of terms in the CIRC Thesaurus; proposed additions shall be limited to terms whose possible inclusion in the CIRC Thesaurus is warranted by sufficient research. The suggested term must satisfy the following conditions: it must express a unique concept (not already expressed in the CIRC Thesaurus by some other term); the concept expressed by the term must be clearly within the area of interest of the S&T intelligence community, and there must be a substantiated demand for the concept.

The originator will complete items 1-5, retain ORIGINATOR SUSPENSE COPY, and forward the remaining three copies to the CIRC Lexicography Section through the designated quality control point (TDBAC-2 for all indexers, TDBIR for all CIRC users). The CIRC Lexicography Section will complete items 6-10, retain LEXICOGRAPHY FILE COPY, and forward remaining two copies (ORIGINATOR FILE COPY and INFORMATION COPY).

NAME/ORG/DATE	Originator's name, organization/symbol, date this form completed.
TERM	Term, descriptor topic tag requiring lexicographic attention/action.
DEFINITION/ BROADER TERM	Description of concept expressed by the term/broader or more generic term under which term in item 2 might be hierarchically structured.
CLASSIFICATION	Level and basis/source (e.g. document, analyst) authorizing/assigning this classification.
ACTION REQUESTED	Proposed addition of new term; deletion/modification to existing term already contained in the CIRC Thesaurus.

Fig. A3 Term Control Form Instructions

31 July 1969

TM-WD-(L)-321/000/00

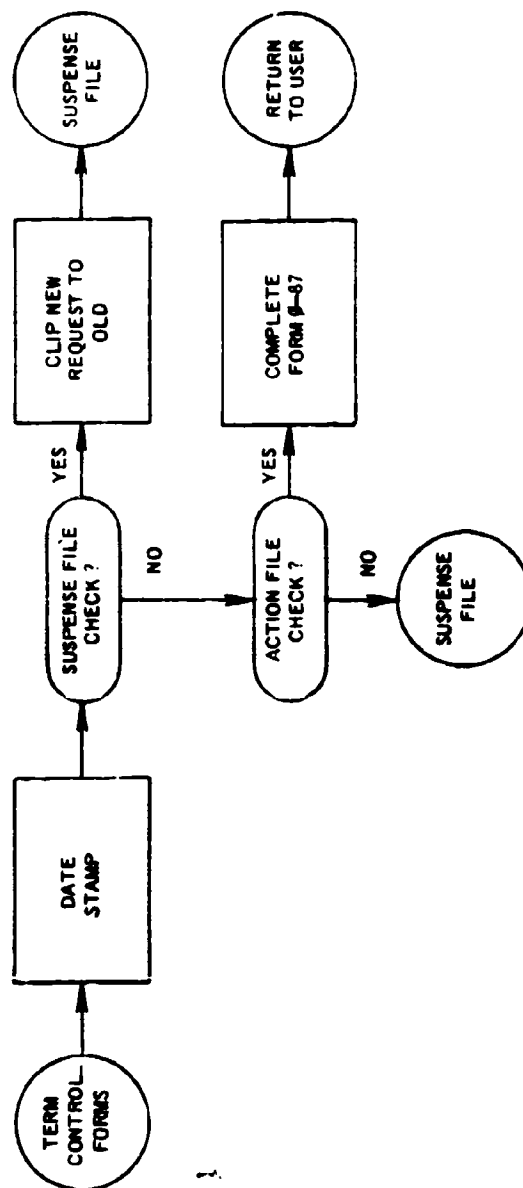


Fig. A4 Duplication Check of Term Control Form

Changes initiated within the lexicographic section are recorded only on the yellow copy of the form. These forms are placed in the suspense file until they are processed.

The present policy is to maintain the suspense file for at least six months before processing the requests into machine-readable form necessary for a vocabulary update. The delay was caused by the low priority given to machine processing for maintaining an up-to-date vocabulary. Production and statistics dominate the operational environment within FTD. A candidate term of the suspense file must have the following characteristics:

1. Uniqueness relative to the other terms in the vocabulary.
2. Application to an accepted interest area.
3. Responsiveness to a legitimate demand.

When a term is rejected by the lexicographic section item six of the Term Control Form is completed. One copy is retained in the lexicographic action file; and the remaining copies are returned to the originators.

Item six is also completed for candidate terms which are accepted; copies 1 and 3 are returned to the originating agency. The lexicography copy is keypunched before being retired to the action file. Figure A5 represents the general decision making process for disposition of terms in the suspense file.

What has been described is the procedure now used within the lexicography section for input, processing, and final disposition of requested term changes or additions to the CIRC thesaurus. Next will be discussed the means presently available to users and indexers for access to the most current and comprehensive version of the CIRC thesaurus.

#### VOCABULARY CHANGE ANNOUNCEMENTS

A most important aspect of lexicographic responsibility is the timely announcement to the user group of all additions and modifications to the system vocabulary. Maximum return can be expected if the changes are displayed already assimilated into the groups and hierarchical structure of the Thesaurus. At present, a revised edition of the CIRC Thesaurus is published annually. This is a three-volume publication that displays the vocabulary in three formats, a structured vocabulary (VOCSS), a permuted index, and an alphabetized vocabulary (THESA). The high cost of producing the thesaurus prevents more frequent publication and distribution.

At six-month intervals, an updated permuted index is produced and distributed. This publication is regarded by users as the most useful of the computer-produced tools for indexing and retrieval. In the past, typewritten lists of all changes and modifications to the vocabulary, with the changes keyed to the proper pages of the permuted index were furnished periodically. Reaction of the user groups to this method of announcement was extremely negative because of the need to post changes manually on the latest printed formats in order to maintain a current and comprehensive display of the vocabulary. In addition, term additions to profiles are reported back to the requestor.



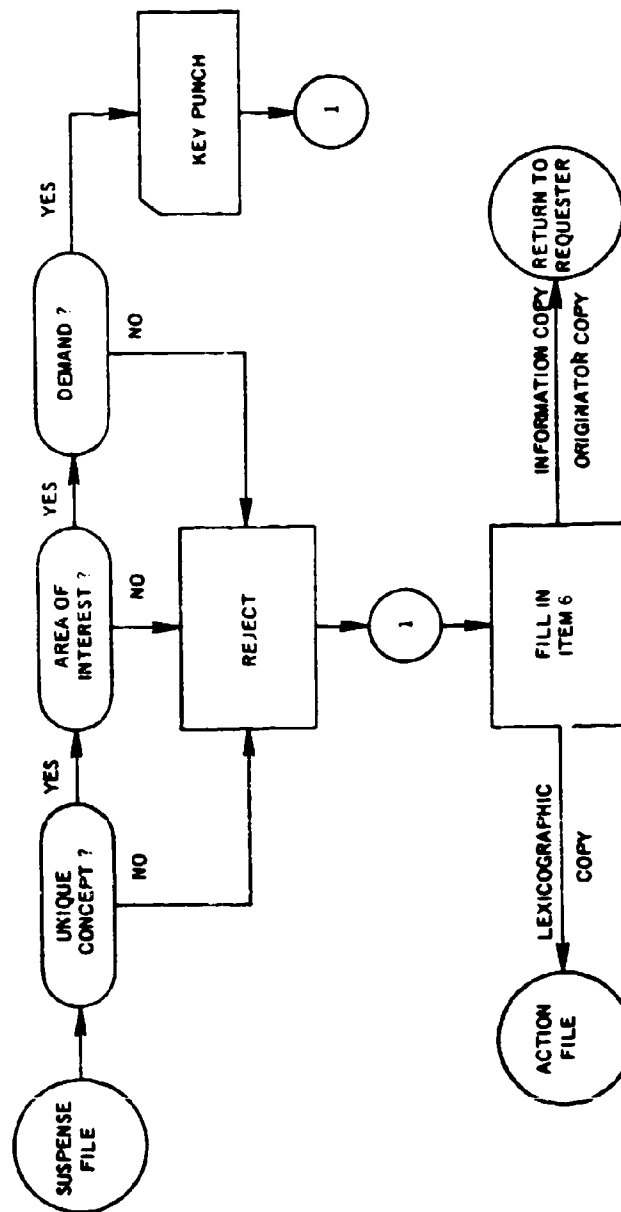


Fig. A5 Suspense File Processing

Other tools produced by the computer support system are geared to the lexicographic group only. Listings showing over-posted terms, frequency counts, rejected terms, unofficial nomenclature, profile information, etc., are printed on demand or as a result of a programmed update of the vocabulary. These tools are most important to the lexicographer, who, in order to make sound judgments on new candidate terms, must have at his disposal all possible information concerning the existing thesaurus.

#### PROBLEM AREAS

Some problems stemming from the consolidation of the microthesauri into an expanded CIRC vocabulary and the impact of additional users into the present system of vocabulary control and update should be anticipated. It is difficult to determine, at this point, what influence the new equipment configuration will have on existing problem areas. The effect could be a complete revision of system operation or, at the very least, a shift in policy from undue concern with production figures to a more realistic proportion of priority given to system maintenance.

The execution of this contract will reduce the vocabulary maintenance backlog. For instance, a major function of the lexicographic section is thesaurus review. In view of the insufficient staff and the lack of computer support at FTD, this would be massive undertaking. Preliminary estimates of new unique terms to be added to the thesaurus were as high as 10,000 terms.<sup>1</sup> This would be a monumental task if it were assigned to the lexicographic section on top of normal work loads. Under the provisions of the contract, however, a group of experienced lexicographers has been assembled to work with FTD in insuring the validity and proper cross referencing of each term in each microthesaurus. This effort should uncover and correct many of the discrepancies which presently exist in the thesaurus.

The areas of concern to this study are those most apt to endanger a simple transition to an expanded CIRC and the introduction of new S&T intelligence activities to the CIRC system. The following points are considered critical to FTD in its responsibility as the Executive Agency of DIA.

1. The lack of proper guidelines and clearly defined standard operating procedures for allowing new activities to enter the CIRC system.
2. Insufficient manning levels within the lexicography office to maintain the vocabulary effectively.
3. Insufficient support from the computer facility, which precludes thesaurus updates on a monthly basis and the production of vocabulary change announcements. This lack of operational support to the quality

---

The actual growth by about 3800 terms was below estimates because many candidate terms failed to meet the criteria of being conceptually unique, and their introduction would have impaired rather than aided utility.

control function is a prime factor in preventing the Thesaurus from becoming truly current and comprehensive.

#### RECOMMENDATIONS

Of primary concern is the need at every new activity for CIRC system documentation, operating procedures for interaction with the system, and technical advice as to the creation of a workable in-house CIRC user unit. In order to meet these needs, it is recommended that:

1. Each activity agency create an information specialist desk which will be the quality control point for all requests from analysts to the CIRC system.
2. The agency information specialist will also be responsible for quality control of all in-house indexing in order to improve its efficacy and consistency with that of other activities.
3. The information specialist act as the single-point liaison for all communication between the FTD lexicographic section and the participating unit.
4. The information specialist shall maintain an action and suspense file on all Term Control Forms originating from his organization.
5. The FTD lexicographic section devise a training program for the information specialist of each new participating agency; such program to include a thorough briefing on system operation and system responsibilities.
6. The FTD lexicographic section prepare a package of guidelines and system documentation designed to orient new activities to their responsibilities in preparing themselves to interact with the CIRC system.
7. The FTD lexicographic section prepare the necessary justification for computer support which will allow for monthly updates of the CIRC thesaurus.
8. The FTD lexicographic section prepare for distribution to the users the necessary support to allow a monthly supplement of thesaurus changes to be produced in the form of a cumulative permuted index of changes, month by month, until the semi-annual publication of the complete permuted index is released.
9. An investigation be undertaken of the application of on-line methods, particularly operating data management systems, to lexicographic functions. (See Section II for further information).

31 July 1969

TM-WD-(L)-321/000/00

10. In order to properly carry out the responsibilities of liaison, instruction and vocabulary control, the FTD lexicographic section be increased permanently to:

1 Supervisor	(1939)
1 Secretary	(GS/5)
2 Lexicographers, Sr.	(1156)
2 Lexicographers	(1373)
2 Clerks	(GS4/5)

## APPENDIX I

## STUDY AND ANALYSIS OF MAINTENANCE PROCEDURES IN CIRC OPERATIONS

## SECTION II

## AUTOLEX

## BACKGROUND

Since the decision was made that the Centralized Information Reference and Control (CIRC) system at the Foreign Technology Division of AFSC was to be the basis for a DOD wide S&T intelligence information processing system, operational capabilities have been improved to enhance the services to the whole intelligence community. A case in point is the DIA-sponsored experiment allowing analysts throughout the intelligence community direct access to a CIRC-like data base through remote teletype consoles. The CIRC On-Line Experiment (COLEX) proved successful and resulted in DIA approval to go operational with the CIRC On-Line (CIRCOL) retrieval system.

The COLEX experience also underlined a basic requirement for a current and comprehensive thesaurus reflecting all terms and their relationships required by each organization and so structured as to satisfy the needs of indexing, retrieval, and dissemination throughout the intelligence community. To meet this requirement, DIA directed an effort to:

1. Create a computerized scientific and technical intelligence thesaurus which would represent the combined interests of all S&T intelligence organizations.
2. Provide a means of systematic updating and control of terminology in this thesaurus.

Persuant to the second goal, there are indications that other means of operational support of the thesaurus should be investigated. This paper describes an exploration of the merits of automating lexicographic (AUTOLEX) functions within the working environment of a Time-Shared Data Management System (TDMS.)

## SCOPE OF THE PROJECT

The AUTOLEX project was designed to provide operational data on a direct-access capability for lexicographic functions. The project was concerned specifically with applying operational direct-access techniques to a large machine-file thesaurus for the purpose of update and modification.

In the interest of economy, the already operational TDMS was chosen as the test vehicle and a test data base was created which resembled as closely as possible the CIRC thesaurus master file in subject-area order. The test data base contained only terminology in basic sciences so that no compromise of sponsor or parochial interests could be implied.

## TDMS ENVIRONMENT

The AUTOLEX project was envisioned as requiring a dialogue capability (through CRT or console) that would permit a lexicographer to query, review, modify, delete, and update the files. In addition, listings in specified formats could be obtained. However, for the purposes of the test, a minimum requirement of scanning, selecting, and updating Official Terms was decided upon. The TDMS system, though incomplete at the time, had the computer programs necessary to meet the minimum test requirements.

## DATA BASE DESCRIPTION

The first ten subject areas (representing about 1300 Official Terms) of the CIRC thesaurus were chosen for the test data base. A fairly simple group of data elements was selected and recorded in a format required by TDMS. More sophisticated test data would not have contributed measurably to the realism of the simulation.

The following elements were chosen:

<u>Element Name</u>	<u>Description</u>
1. TERM DESIGNATOR	OT, ONT, SA, SN
2. SUBJECT CODE	1 through 56
3. TERM LEVEL	1 through 5
4. COSATI CODE	Four digit code from TEST or other COSATI based sources
5. OFFICIAL TERM	
6. TERM CODE	Numerical position in FTD file
7. COSATI LEVEL	1 through 5
8. COSATI TERM CODE	Numerical position in COSATI file
9. FREQUENCY COUNT	NUMBER OF POSTINGS
10. PROFILE DATA	
11. USER CODE	ACTUAL UNIT DESIGNATORS
12. AGENCY DATA	
13. AGENCY CODE	TERM IS UNIQUE TO....

## TDMS COMPONENTS

The following operations were utilized during the test phase:

1. The Define operation, which is necessary to describe the data elements to the system (Fig. A6).
2. The Generate (Load) which is used to create a TDMS data base or to add entries to an existing data base (Fig. A7).
3. The Query operation, which permits the retrieval and printing of selected data from a specified data base. A limited capability for calculations is provided (Fig. A8).
4. The Update operation, which allows changes to existing data values and the addition or deletion of data values or entire logical entries to and from a data base (Fig. A9).
5. The Compose operation, which is designed to permit a user to describe the format of a report for subsequent generation (Fig. A-10).

## PROJECT DIARY

A duplicate deck was produced from the existing thesaurus master file. The elements represented were subject code, level code, and official term. All additional elements required were arbitrarily created and recorded on work-sheets for keypunching on the existing cards.

On the advice of the TDMS maintenance staff at Falls Church, it was decided to run the Define and Load operations at Santa Monica to take advantage of the experience of the staff, to eliminate line charges for Falls Church access, and to save time in the creation of the data base.

The Define operation gave no problems; however, the Load operation caused some difficulties. Two days were spent in correcting data base errors brought about by misinterpretation of TDMS user guides. Then, with the data base in fairly good order, the next four days were spent in trying to complete the Load operation under the time sharing system. As it turned out, the mean-time-to-failure of the time sharing system (with 10 to 20 users) was less than the time required to complete the Load operation. This necessitated the acquiring of the computer on a special request to operate with no other users. Under these conditions the load time took twenty minutes.

The remaining TDMS components were tested at Falls Church via the remote teletype hook up with the Santa Monica computer facility. The Query and Update components were run exhaustively, but little time was realized on the Compose component because of its state of completion at the time.

## CONCLUSIONS

For a number of reasons, the AUTOLEX project was not given a true test. In a technical sense, the environment lacked the crucial interaction with the production side of the CIRC operation. The only proof of success in updating the

31 July 1969

IM-WD-(L)-321/000/00

```
/login msdhourly 30
LOGGED IN 10:12 10/30/68
/define
PGM.STARTED
DEFINE VERSION 5.2 IS OPERATING
ENTER 'DEFINE OR REDEFINE': define
DATA BASE NAME IS: alex
ENTRY TERMINATOR IS: end
ENTER 'TTY' OR INPUT FILE IDENTIFIER: tty
BEGIN INPUTS
1 term designator (name)
NEXT
2 subject code (number) values are 01...56
NEXT
3 term level (number values are 1...5
NEXT
4 cosati code (number) format is 0999 values are 0100...2204
FORMAT ACCEPTED.
NEXT
5 official term (name)
NEXT
6 term code (number)
NEXT
7 cosati level (number) values are 1...5
NEXT
8 cosati term (number)
NEXT
9 frequency count (number) values are 0...3999
NEXT
10 profile data (repeating group)
NEXT
11 user code (name in 10)
NEXT
12 agency data (repeating group)
NEXT
13 agency code (name in 12)
OUTPUT FILE IDENTIFIER IS: dalex
FILE DALEX, 2314 VOL. 9001 OPENED ON AO
DEFINE OPERATION SUCCESSFULLY COMPLETED
DEFINE VERSION 5.2 IS OPERATING
ENTER 'DEFINE OR REDEFINE': / quit define
PGM.QUIT
```

(Note - lower case text were input at console; upper case text was system response)

Fig. A6 AUTOLEX Define Example



31 July 1969

TM-WD-(L)-321/000/00

```
/generate
PGM.STARTED
GENERATE VERSION 5.2 IS OPERATING
ENTER OUTPUT FILE: autolex v40346
NEW/ADDON/RESTART: new
2314 VOL 0346 TO BE MNTED ON A6
FILE AUTOLEX, 2314 VOL. 0346 OPENED ON A6
ENTER DESCRIPTION FILE: dalex v40346
ENTER 'TTY' OR INPUT DATA FILES: autape v95983 S2
THE DATA BASE NAME IS 'ALEX'.
TAPE VOL 5983 TO BE MNTED ON CA
NEXT
trace 100
NEXT
run
FILE LCONELEX, 2314 VOL. 9003 OPENED ON A2
FILE LFAILLEX, 2314 VOL. 9001 OPENED ON A0
FILE CFINDLEX, 2314 VOL. 9002 OPENED ON A1
FILE LNAMELEX, 2314 VOL. 9003 OPENED ON A2
'100' ENTRIES PROCESSED
'200' ENTRIES PROCESSED
'300' ENTRIES PROCESSED
'400' ENTRIES PROCESSED
'500' ENTRIES PROCESSED
'600' ENTRIES PROCESSED
'700' ENTRIES PROCESSED
'800' ENTRIES PROCESSED
GENERATE OPERATIONS SUCCESSFULLY COMPLETED
```

Fig. A7 AUTOLEX Generate Example

31 July 1969

TM-WD-(L)-321/000/00

PGM. STARTED

QUERY VERSION 5.2 IS OPERATING

ENTER DATA BASE FILE IDENTIFIER: autolex w40346

FILE VOL 0346 TO BE MNTED ON B4

TEXT

Describe

C1 TERM DESIGNATOR (NAME)  
C2 SUBJECT CODE (NUMBER) VALUES ARE 01...56  
C3 TERM CODE (NUMBER) VALUES ARE 1...5  
C4 COSATI CODE (NUMBER) FORMAT IS 0999 VALUES ARE 0100....2204  
C5 OFFICIAL TERM (NAME)  
C6 TERM CODE (NUMBER)  
C7 COSATI LEVEL (NUMBER) VALUES ARE 1...5  
C8 COSATI TERM (NUMBER)  
C9 FREQUENCY COUNT (NUMBER)  
C10 PROFILE DATA (REPEATING GROUP)  
C11 USER CODE (NAME IN 10)  
C12 AGENCY DATA (REPEATING GROUP)  
C13 AGENCY CODE (NAME IN 12)

---  
/show c6  
----

1	V1	1
1	V111	222.0

SEARCH: \$

---  
print entry where c6 gq 150  
----

C1=OT C2=4 C3=1 C4=502 C5=LIBRARY C6=222.0 (C12)C13=FTD (C12)C13=STIC  
----

print entry where c9 gq 700  
----

C1=OT C2=4 C3=1 C4=509 C5=SCIENTIFIC PERSONNEL C6=762 (C10)C11=AMD  
(C12)C13=FTD (C12)C13=STIC

C1=OT C2=4 C3=2 C4=501 C5=TECHNICAL ASSISTANCE C6=92 714 (C10)C11=DPCE/H  
(C12)C13=FTD (C12)C13=STIC

---  
/quit query  
PGM.QUIT

Fig. 8 AUTOLEX Query Example

31 July 1969

TM-WD-(L)-321/000/00

```
/update
PGM.STARTED
UPDATE VERSION 5.2 IS OPERATING
ENTER DATE BASE FILE IDENTIFIER: autolex v40346
-----
DATA BASE NAME IS 'ALEX', MOD 14, DATE 12/10/68
USE IT, 'Y/N': y
NEXT
print entry where c9 eq 765
-----
C1=OT C2=4 C3=1 C4=509 C5=SCIENTIFIC PERSONNEL
C6=86 C9=765 (C10)C11=AMD (C12)C13=FTD (C12)C13=STIC
-----
add profile data, user code = jjb where c9 eq 765
-----
'1' ENTRIES, '1' OCCURRENCES SELECTED, 'EX/MORE/S':
ex
NE
print entry where c9 eq 765
-----
C1=OT C2=4 C3=1 C4=509 C5=SCIENTIFIC PERSONNEL C6=86 C9=765 (C10)C11=JJB
(C10)C11=AMD (C12)C13=FTD (C12)C13=STIC
-----
add profile data, user code=jjb where c9 eq 714
-----
'1' ENTRIES, '1' OCCURRENCES SELECTED, 'EX/MORE/S':
ex
NEXT
print entry where c9 eq 714
-----
C1=OT C2=2 C4=501 C5=TECHNICAL ASSISTANCE C6=92 C9=714 (C10)C11=JJB
(C10)C11=DPCE/H (C12)C13=FTD (C12)C13=STIC
-----
/quit update
PGM.QUIT
```

Fig. A9 AUTOLEX Update Example

31 July 1969

TM-WD-(L)-321/000/00

```
/compose
PGM.STARTED
COMPOSE VERSION 5.2 IS OPERATING
ENTER DATA BASE FILE IDENTIFIER: autolex v40346
ENTER COMPOSE FILE IDENTIFIER OR 'NONE':
none
NEXT
qualify entry where c9 gq 500
NEXT
title is unclassified
'T1' ACCEPTED

heading is subj area, level, term, freq cnt
'H1' ACCEPTED
NEXT
content is subject code, term level, official term, c9
'C1' ACCEPTED
NEXT
PUT TOP t1, h1
NEXT
run
REPORT GENERATION HAS BEGUN
DATA BASE NAME IS... ALEX
DATE OF REPORT GENERATION IS... 12/11/68
```

UNCLASSIFIED			
SUBJ AREA	LEVEL	TERM	FREQ CNT
4	1	SCIENTIFIC PERSONNEL	765
4	1	SCIENTIFIC RELATION	615
4	2	TECHNICAL ASSISTANCE	714
4	3	FOREIGN TECHNICAL ASSISTANCE	502
4	1	SCIENTIFIC CONFERENCE	566

```
---
/quit compose
PGM.QUIT
```

Fig. A10 AUTOLEX Compose Example

thesaurus would be the successful operation of the files by retrieval and disseminating functions in real time with usual volumes. Also, the limitations of TDMS (it was not created as a lexicographic tool) prevented a realistic demonstration of providing direct access to a machine-file thesaurus. For instance, each TDMS operation must be "called in" and "quit" before any other can be brought to play, resulting in an undesirable discontinuity.

The number of elements chosen for the entry description was reduced to save time in creating the data base, and was not analogous to that encountered in an operational situation. Despite these difficulties, the test results justify continuance of the effort. The results show that on-line processing will improve existing thesaurus maintenance capabilities in the following ways:

1. Terms can be added to the thesaurus at any time. Updatings more frequently than on the present six month cycle would mean that the vocabulary would be "correct" at any given time. Crash programs to meet deadlines for updating runs would never be necessary.
2. The time lag would be reduced between the submission of suggested terms by users and the availability of those terms to indexers and searchers.
3. The complex filing system could be simplified, freeing personnel in the lexicography office for more productive tasks.

#### RECOMMENDATIONS

The following points should be considered in developing test criteria for a useful and practical vocabulary control and maintenance system.

1. The test should be operated under a system of direct access much like the ORBIT/360. This would require special programs tailored to lexicographic functions.
2. The test should be operated by lexicographic personnel in order to arrive at the best possible combination of man-machine specifications.
3. Consideration should be given to all possible types of interacting devices, such as display units, consoles, and teletypes. As a special part of the test, displays could be given to indexers and analysts for the purposes of screening and selection (but not updating) of the most current thesaurus.
4. Special consideration should be given to the "Compose" or Report Generator function of an AUTOLEX project. This could very well be the most important feature of such a maintenance system. The ability to change a report format without outside programming support, would provide a degree of flexibility to vocabulary maintenance operations that would extend present capabilities significantly. For example, studies could be conducted of the indexing and retrieval effectiveness of terms; term utilization in profiles; term inter-relationships based on profiles, indexing, and searching; and experimental term display formats.

## APPENDIX II

## COMMENTS ON THE CIRC THESAURUS

The expansion of the vocabulary in the subject fields of interest to MID, MIO, STIC, and DIAST-2C brought about specific recommendations for changes in the organization of the CIRC Thesaurus dated 31 December 1968. In addition, recommendations were made for the deletion of terms that were believed to be redundant, overly specific, insufficiently defined, or otherwise unsuitable for inclusion in the DOD S&T Intelligence Thesaurus. Apart from these recommendations, a conscious effort was maintained to adhere to the principles of term selection, term construction, and the display of term relationships that had been set forth in the CIRC Lexicographic Guide (FB-DA(L)-150/020/02) and that were evidenced in the CIRC Thesaurus. In this sense, the compilation of the S&T Thesaurus was not regarded as wholesale revision of the CIRC Thesaurus.

As the work progressed, it became clear that the CIRC Thesaurus philosophy exerts certain constraints upon vocabulary development, particularly upon vocabulary display. The purpose of this discussion is to set forth in some detail the nature of these constraints and to recommend ways in which the thesaurus format can be altered to provide an indexing and retrieval tool of greater flexibility and utility. This should not be considered a critique of the existing system or of the personnel involved. Rather, it is an exploration of ways to increase the capabilities of CIRC, to amplify the explanation of the CIRC philosophy in the Lexicographic Guide and in the thesaurus introduction, and to set forth some considerations for other related vocabulary efforts. The latter could be particularly significant in any ensuing thesaurus compilation projects.

## TERM SELECTION

The key to compiling a thesaurus is the determination of what concepts are to be represented. It is essential that considerations of format or term display not be allowed to obscure the importance of content. Once a concept has been identified, there follows the determination of:

1. The term that best connotes that concept.
2. The exact construction of the term.
3. The relationship(s) of the term to other terms in the vocabulary.

The most recent CIRC Thesaurus contained many terms for which meanings could not be ascertained, several cases of apparent synonymy between Official Terms, and a great many specific terms connoting concepts that could be represented by a combination of more general terms. Moreover, many inconsistencies with respect to term construction were noted. It was practical to rectify only the most serious of these discrepancies while fulfilling the provisions of the contract, but it seems likely that the difficulties brought about by these situations will be compounded as the vocabulary grows in size and scope.

The exact causes for these discrepancies are not fully understood, but it seems likely that several factors contributed. Among them are:

1. The inclusion in the original CIRC vocabulary of poorly defined or unnecessary terms from the Intelligence Subject Code (ISC).
2. Workloads and publication deadlines at FTD that prohibited adequate research and consideration of candidate terms.
3. An organizational structure of FTD which treats lexicography as a non-technical function subordinate to indexing and analysis.
4. An aversion on the part of FTD lexicographers to terms having high posting densities or rapid increases in posting density.
5. The tendency of the update programs to militate against drastic changes in the thesaurus structure and content.

At the time the original vocabulary was organized, there was a requirement to adhere to the structure of the ISC. This gave rise to some terms that were coined to cover an area or range of ISC numbers but were not expressions of definable concepts. For other cases, a heterogeneous collection of materials or items was covered by a generic term which was not given a Scope Note. Other broad terms were incorporated for structuring utility, but had no caveat to encourage the use of more specific Official Terms. As a result, many overly general, poorly defined or vague terms became ingrained in the vocabulary and have set poor precedents for term selection and construction. Examples are: SYNTHETIC MATERIAL, NONSTRUCTURAL MINERAL PRODUCT, INTERMOLECULAR FORCE, MATTER STRUCTURE and TRANSPORTATION STATUS.

The lexicographers at FTD are permitted to exercise relatively little discretion with respect to the evaluation of terms that are submitted by indexers, analysts, and CIRC users. This, coupled with heavy workloads and tight publication schedules, is believed to have caused many terms of dubious indexing utility to be added to the thesaurus over a period of years. In addition, the FTD policy is to retain any term that appears in any CIRC profile whether or not it has ever actually been used in indexing. This obviously makes editing the thesaurus difficult.

For some time the practice at FTD has been to prohibit the use of terms which have a frequency of posting that exceeds a certain predetermined level. The usual approach has been to lower the frequency of posting by the synthesis of many very specific terms that connote various aspects of the general concept. This has produced a proliferation of specific terms in a few areas, notably materials, chemistry, and metallurgy, that may unnecessarily contribute to difficulties of vocabulary maintenance. For example, of some 70 terms relating to copper, many are extremely specific and connote concepts that could be represented by combinations of valid terms that are more general.

These difficulties are compounded by the fact that the existing maintenance programs are designed to facilitate the adding of new terms to the thesaurus. Since reorganization of terms can only be accomplished by two updating runs, this has, not been undertaken too frequently. (This problem was dealt with in the current revision effort, but it undoubtedly accounts for many existing discrepancies.)

With regard to the construction of terms, the existing procedures make several specific provisions -- some of which seem strangely arbitrary -- regarding term construction, but the thesaurus contains a great many instances in which these provisions have not been observed. For example, the rule against plural forms properly provides that terms connoting scientific disciplines are not to be considered plurals; the thesaurus includes seven terms containing the word "geophysic" yet includes still other inconsistent patterns as illustrated by terms such as MATHEMATIC CONFERENCE and MATHEMATICS INSTITUTE. There is an apparently arbitrary rule against adjectives ending in "-al" which has resulted in such awkward constructions as STATISTIC ANALYSIS and ELECTRIC ENGINEER. At the same time, inconsistencies such as STATISTICAL THERMODYNAMICS and GLASS ELECTRICAL CONDUCTIVITY have found their way into the thesaurus. The overall effect of these unsuccessful efforts to systematize term construction has been to create terms that are unnecessarily awkward and synthetic. This might be better stated that the Natural Language has become quite unnatural.

It should be noted that there were times when the technical judgments of the project were tried or bent to maintain a consistent pattern in term construction. The bending seemed to generally conflict with natural language as it appears in the text of documents being processed or in normally expected expressions for dissemination and retrieval. These concerns stem from the realization that the end use of the thesaurus is a tool, and user requirements for concept labelling should take precedence over lexicographic preferences for terms.

#### THESAURUS FORMAT

Generally, the thesaurus format is quite useful and compares favorably with that of other indexing vocabularies. There are, however, certain constraints inherent in the format which, if they cannot be circumvented, should be dealt with much more candidly in the Lexicographic Guide and in the thesaurus introductory material.

#### Alphabetized Vocabulary

The alphabetized vocabulary should be considered the most important indexing and retrieval tool. Since this section lists terms alphabetically, it should be useful to users having any degree of familiarity with the vocabulary. Moreover, the most complete information about the terms and their disposition in the other displays is given here.

##### 1. Scope Notes

Scope Notes have been misused in a few cases in the CIRC thesaurus; i.e., they have been used to supply dictionary definitions, rather than to explain ambiguous or closely overlapping terms. Of more importance are the many terms that convey no obvious meaning and for which no Scope Notes have been provided.

##### 2. Synonyms

There are relatively few actual synonyms among technical terms, but, there are many instances in which the concepts represented by sets



of terms overlap to such an extent that a single valid indexing concept is represented. There are even more instances in which specific concepts can be indexed by a combination of more general terms. When properly used, the device of designating as "synonyms" certain sets of terms will enhance indexing and retrieval consistency and will prevent the detrimental proliferation of terms.

FTD has recognized and rectified many instances of poorly chosen synonyms, but several still exist. In addition, there are many cases in which terms overlap to such an extent that they should be considered synonyms for indexing and retrieval purposes.

### 3. See Also references

There has been no consistent use of See Also references. The importance of these references, particularly in view of the constraints imposed by the assignment of terms to one subject area and one hierarchy has been underestimated. Adequate criteria for the establishment of See Also references have not been developed.

### Subject - Structured Vocabulary

The subject structure was patterned closely after the ISC. This was a requirement in the original CIRC vocabulary and had the advantage of helping to insure complete coverage of the scientific portion of the ISC. It had the disadvantage of imposing upon the CIRC thesaurus a subject structure that was in some respects outmoded and unduly arbitrary. The questionable utility of some of the terminology of this structure was noted above. Not surprisingly, the results of FTD building upon this base has resulted in a subject categorization and hierarchical structure that confuses new users who are unaware of the original criteria.

### 1. Vocabulary groupings

The 56 vocabulary groupings include the major fields of science and engineering, a few categories relating to military science and technology, and a few groups of miscellaneous content. On the surface, there appear to be no significant shortcomings in this categorization. However, some problems have arisen in the assignment of the thesaurus terms to these groups. These problems stem from:

- a. The inherent arbitrariness and rigidity of categorization schemes.
- b. The assignment of each term to only one category.
- c. Inconsistent interpretation of the scope of some categories.
- d. Failure in some cases to determine and to accommodate user requirements.

The subject structured-vocabulary must be recognized as only an adjunct to the alphabetized vocabulary. It is a device for a display of certain term relationships and its usefulness is in direct proportion to the degree of familiarity that a given user has with the thesaurus. The arbitrary nature of subject categorizations and the arbitrary decisions that are required to assign new terms to categories make this display relatively useless to anyone unfamiliar with its arrangement and content. On the other hand, frequent thesaurus users can learn to use the display effectively, no matter what the arrangement.

The main objective of a display of this kind is to provide thesaurus users, i.e. indexers and searchers, with a quick reference to a subset of the total vocabulary. To be useful, the subset should:

- a. Encompass subject matter that will correspond in some way to that of items being indexed or searches being conducted.
- b. Represent a comprehensive and coherent treatment of that subject matter.
- c. Present a display of a size that can be scanned quickly and easily.

In subdividing a large heterogeneous vocabulary, such as that of CIRC, the isolation of certain more obvious, desirable, or coherent categories inevitably creates a residue of miscellaneous terms that give rise to some rather awkward categories. These categories of miscellaneous terms need not unduly influence, or detract from, the more desirable coherent categories. For instance miscellaneous categories, such as Chemical Products, Scientific Instruments and Other Products and Equipment should not attract sets of terms which are closely or exclusively oriented to another subject area. Some examples of "attracted" terms which have more logical alternate locations are: DRUGS which appears in the subject area Chemical Products, but from some points of view might very well be assigned to Medical Sciences; METEOROLOGICAL INSTRUMENT which appears in Scientific Instruments, but might well be assigned to Meteorology; and PHOTOGRAPHIC EQUIPMENT, which by its appearance in Other Products and Equipment, is separated from many closely related photography concepts that are found in Optics.

When terms can be assigned to only one category, subjective judgments regarding the disposition of terms are required. A given category should be judged primarily on the basis of its value as a display device in relation to the vocabulary as a whole, and only incidentally in relation to other categories. An overall rationale and certain guidelines should be employed to facilitate updating, but consistency per se is not of primary importance. The goal should be to satisfy the requirements of a majority of thesaurus users.

## 2. Hierarchical relationships

The hierarchical relationships displayed in the subject-structured section and, by means of BT-NT cross references, in the alphabetized vocabulary are quite valuable in that they provide a degree of organization to the vocabulary as a whole and implied definitions to individual terms. Searches conducted by means of the hierarchies

can contribute to the effectiveness of the retrieval system.

The hierarchies in the CIRC Thesaurus have been developed from various points of view and with varying degrees of consistency. Serious constraints are imposed by the assignment of terms to only one hierarchy and from the limitation on the number of levels a hierarchy may contain.

A simple example of inconsistency in the development of hierarchies in the CIRC thesaurus is the treatment of the terms ASTRONOMY and ASTROPHYSICS in Subject Area 06. Both terms are level one, implying that they are considered separate fields of study, but the term ASTROPHYSICS CONFERENCE appears subordinate to ASTRONOMIC CONFERENCE, implying that astrophysics is a branch of astronomy. Further, PRACTICAL ASTRONOMY and DESCRIPTIVE ASTRONOMY appear in separate hierarchies from ASTRONOMY, but ASTRONOMIC GEODESICS is subordinate to ASTRONOMY. ASTRONOMIC OBSERVATION is subordinate to ASTRONOMIC GEODESICS, but presumably could apply equally well to practical or descriptive astronomy.

The difficulty that arises from assigning terms to only one hierarchy is shown by the terms subordinate to PHOTOGRAPHIC ASTRONOMY. These terms, LUNAR PHOTOGRAPHY, PLANETARY PHOTOGRAPHY, SOLAR PHOTOGRAPHY and STELLAR PHOTOGRAPHY, would be equally appropriate as members of the PHOTOGRAPHY hierarchy that appears in field 10. A hierarchical search of the term PHOTOGRAPHY would yield only some of the appropriate references in the file.

#### PERMUTED VOCABULARY

This is an excellent means of displaying the thesaurus vocabulary. It can be seen that consistent construction of analogous terms is important since similarities or coincidences in the words that form the terms provide the basis for showing one kind of relationship among terms.

#### RECOMMENDATIONS

1. A review should be made of the lexicography function at FTD with a view toward elevating its relative position to one of responsibility for the technical content as well as other aspects of the vocabulary.
2. A review should be made of the thesaurus format with emphasis on:
  - a. Recognition of the alphabetized section as the most important part of the thesaurus and the other displays as adjuncts to it.
  - b. An objective assessment of the validity of the existing subject categorization and its responsiveness to user requirements.
  - c. A feasibility study of multiple hierarchical assignment of terms.

31 July 1969

TM-WD-(L)-321/000/00

- d. The introduction of, instead of the SY reference for synonyms, an indexing instruction that would prescribe the use of one, two, or perhaps more terms to index a given concept.
  - e. The establishment and implementation of criteria for developing See Also references among terms.
3. A reevaluation of thesaurus terminology should be conducted based on the utility of the terms in indexing and retrieval. Terms that have not been used should be deleted. Terms that have been used very infrequently in the past three or so years should be considered likely candidates for deletion or for being made synonyms of valid terms. Terms for which no meaning is obvious or readily available, as in reference work, should be investigated. Those that are valid, e.g. in terms of reflecting usage terminology and unique concepts, should be given Scope Notes. Others should be deleted or made synonyms of valid terms.
  4. Practical conventions for term construction should be established and applied consistently.
  5. The problem of "overposted" terms should be reassessed. Valid terms that have been posted so frequently that they are no longer useful in retrieval should be so designated in the thesaurus to discourage their use when suitable alternatives are available.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & C		
(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)		
1. ORIGINATING ACTIVITY (Corporate author) System Development Corporation 5720 Columbia Pike Falls Church, VA 22041		25. REPORT SECURITY CLASSIFICATION Unclassified
3. REPORT TITLE  COMPILATION OF DOD SCIENTIFIC AND TECHNICAL INTELLIGENCE THESAURUS		26. GROUP
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Final Report March 1968 - July 1969		
5. AUTHOR(S) (First name, middle initial, last name)  John A. Dovel, Jr.		
6. REPORT DATE October 1969	72. TOTAL NO. OF PAGES 29	76. NO. OF REFS 0
22. CONTRACT OR GRANT NO. F30602-68-C-0252	20. ORIGINATOR'S REPORT NUMBER(S) TM-WD-(L)-321/000/00	
8. PROJECT NO. 9117	20. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) RADC-TR-69-308	
10. DISTRIBUTION STATEMENT This document may be further distributed by any holder only with specific prior approval of DIA (DIAST-1), Wash DC 20315.		
11. SUPPLEMENTARY NOTES		17. SPONSORING MILITARY ACTIVITY Rome Air Development Cntr. (EMPIH) Griffiss Air Force Base, New York 13440
9. ABSTRACT A description is presented of the compilation of a controlled vocabulary for subject indexing scientific and technical information within the DOD intelligence community. The product of the effort was a computer-generated thesaurus comprising some 22,540 terms which corresponds closely to the CIRC Thesaurus (FTD-TM-22-07-69) with regard to format and term construction. The final thesaurus was compiled after preparing four microthesauri for other activities in the intelligence community which were then integrated with the CIRC thesaurus. The DOD S&T Intelligence Thesaurus is divided into three parts: a subject-structured display in which the terms are hierarchically related within 56 broad subject areas; a permuted display in which the single - and multi-word terms are arranged by computer in the order of each word that appears within the terms; and an alphabetical display in which the terms are listed alphabetically with cross references among subject-related or hierarchically related terms and to synonyms where applicable, and with scope notes to indicate the intended usage of ambiguous terms. Appended to the report are discussions of the existing procedures for thesaurus maintenance at CIRC and of some considerations concerning the philosophy of the CIRC thesaurus, with recommendations regarding both. Also appended is a description of a preliminary study of thesaurus maintenance using on-line data processing techniques.		

DD FORM 1473  
1 NOV 65

UNCLASSIFIED

Security Classification

UNCLASSIFIED  
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Thesauri Subject Indexing Information retrieval Technological Intelligence CIRC thesaurus On-line data processing						

UNCLASSIFIED  
Security Classification

APIC--Griffiss AFB NY 28 Nov 69-46